Project Proposal

## Group Members:

Hung Do <hd386@drexel.edu>

Matt McKenna <mm5679@drexel.edu>

Ao Wang <aw3338@drexel.edu>

## Domain and Application

The domain of this project is anime recommendation, a specific recommender system used to help users find anime titles that match their interests, tastes, and needs.

The application of this project is to design and implement an anime recommendation system. This project aims to provide users with a personalized and intriguing anime discovery experience. With this goal in mind, we aim to increase user satisfaction and loyalty to our system while increasing sales and revenue to anime titles that users might not have seen before through growing exposure and visibility.

## Dataset

For our project, we’ll focus on anime datasets found on Kaggle. Both datasets are from the myanimelist.net API and contain information on user preference data. However, the first dataset from 2017 has 73,516 users on 12,294 anime. The other is a much larger compilation from 2020 with 325,772 users and 17,562 anime. Users can add anime to their completed list and give it a rating. The dataset contains two files, the anime.csv and ratings.csv.

For the anime file, the schema’s:

1. anime\_id - myanimelist.net's unique id identifying an anime.
2. name - full name of anime.
3. genre - comma separated list of genres for this anime.
4. type - movie, TV, OVA, etc.
5. episodes - how many episodes in this show. (1 if movie).
6. rating - average rating out of 10 for this anime.
7. members - number of community members that are in this anime's "group".

For the rating’s file, the schema’s:

1. user\_id - non identifiable randomly generated user id.
2. anime\_id - the anime that this user has rated.
3. rating - rating out of 10 this user has assigned (-1 if the user watched it but didn't assign a rating).

Sources:

* <https://www.kaggle.com/datasets/CooperUnion/anime-recommendations-database>
* <https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020>

## Algorithms and Evaluation

Our project will explore three different algorithms to predict users’ explicit feedback. This feedback is given on a 10-star system.

The three algorithms we plan to pursue are:

* Neighborhood-based Collaborative Filtering
* Content-Based Recommendation
* Machine Learning models (Regression and Classification)

We will evaluate the algorithms

* RMSE for regression forecasts
* Confusion Matrix metrics (e.g. Precision) for classification forecasts